Title: THE DEVELOPMENT AND ASSESSMENT OF A SCALE TO MEASURE THE EXPERIENCE OF AN ANOREXIC VOICE IN ANOREXIA NERVOSA

Short title: *DEVELOPING A SCALE TO MEASURE THE ANOREXIC VOICE*

Dr Kay Hampshire*[1], Dr Stephanie Tierney[2], Dr Filippo Varese[3], Professor Gillian Haddock[4], Dr Saeideh Saeidi[5], Dr John RE Fox[6]

[1] Clinical Psychologist, Division of Psychology and Mental Health, University of Manchester, Manchester Academic Health Science Centre, Oxford Road, Manchester, M13 9PL. Greater Manchester Mental Health Trust. Telephone: 07540 294383. Email: kayhampshire700@hotmail.com[1]

[2] Researcher in Evidence Synthesis, Nuffield Department of Primary Care Health Sciences, University of Oxford, Radcliffe Observatory Quarter, Woodstock Road, Oxford, OX2 6GG. Telephone: 01865 289 324. Email: stephanie.tierney@phc.ox.ac.uk

[3] Senior Lecturer in Clinical Psychology, Division of Psychology and Mental Health, University of Manchester, Manchester Academic Health Science Centre, Oxford Road, Manchester, M13 9PL. Greater Manchester Mental Health Foundation NHS trust. Telephone: 0161 3060400. Email: filippo.varese@manchester.ac.uk

[4] Professor of Clinical Psychology, Division of Psychology and Mental Health, University of Manchester, Manchester Academic Health Science Centre, Oxford Road, Manchester, M13 9PL. Greater Manchester Mental Health Foundation NHS trust. Telephone: 0161 3060400. Email: Gillian.haddock@manchester.ac.uk

[5] Head of Clinical Audit and Service Evaluation, Leeds and York Partnership NHS Foundation Trust, Leeds . Telephone: 0113 8559334. Email: s.saeidi@nhs.net

[1] Corresponding author current information: Dr Kay Hampshire, Senior Clinical Psychologist, Sir Sidney Hamburger Unit, Park House, North Manchester General Hospital, M6 5RB. Greater Manchester Mental Health Trust. Telephone: 07540 294383. Email: kayhampshire700@hotmail.com

[6] Consultant Clinical Psychologist / Clinical Director/ Senior Lecturer in Clinical Psychology (honorary). Cardiff University and Prifysgol Caerdydd and Cardiff and Vale NHS board. Email: johnfoxresearch@gmail.com

Conflict of Interest

The authors have no conflict of interest to declare.

Acknowledgments

The authors would like to thank staff and colleagues at the following services for supporting this study: Beating Eating Disorders (Beat), Cardiff and Vale University Health Board, Central and North West London NHS Foundation Trust, Cwm Taff University Health Board, Greater Manchester West NHS Foundation Trust, Leeds and York Partnership NHS Foundation Trust, Manchester Mental Health and Social Care Trust, The National Association of Anorexia and Associated Eating Disorders (ANAD), The National Eating Disorders Association (NEDA), NHS Stockport, Tees, Esk and Wear Valleys NHS Foundation Trust.

Abstract

The anorexic voice (AV) is defined as a critical internal dialogue which has been implicated in the development and maintenance of anorexia nervosa (AN). Systematic research to explore this further requires a valid and reliable measurement tool. This study aimed to develop and assess the validity of the Experience of an Anorexic VoicE Questionnaire (EAVE-Q). EAVE-Q items were developed and checked for face and content validity through cognitive interviews with seven individuals diagnosed with AN. Participants with a diagnosis of AN (N = 148) completed the EAVE-Q, sociodemographic questions and measures of mood and quality of life to assess internal consistency and construct

validity. Forty-nine participants completed the EAVE-Q twice more to assess test-retest reliability. The EAVE-Q had good face and content validity and good acceptability. Principal axis factoring resulted in an 18-item scale organised into five domains with high internal consistency (α = .70 to α = .85). Domains correlated significantly with eating disorder symptoms, psychological distress and quality of life. The EAVE-Q did not discriminate between participants on the basis of body mass index. Test-retest reliability was moderate. Although the factor structure of the EAVE-Q requires replication in other AN samples, the EAVE-Q is the first measure of a critical internal dialogue in AN. It is hoped that it will aid future research to increase understanding of AN and the continued development of person-centred treatments.

Practitioner Points:

- Research has indicated that people with a diagnosis of anorexia nervosa report that they experience an internal 'voice' that gives the eating disorder an identity and drives its severity.
- The relationship between the anorexic voice and the self appears to be crucial in understanding its role in increasing eating disorder severity. There is a lack of an evidence-based measure of the relationship between the self and the 'voice'.
- The EAVE is an evidence-based measure that looks at the relationship between the self and the anorexic voice.

Introduction

Qualitative research in Anorexia Nervosa (AN) has identified the experience of an 'anorexic voice' (AV) (Higbed & Fox, 2010; Tierney & Fox, 2010), hypothesised to be a key psychological component which may contribute to the development and maintenance of AN. The AV is defined as "a critical-internal dialogue (i.e. a second or third commentary which is 'heard'), orientated around shape, weight, eating and their implications for self-worth" (Pugh, 2016, p.1). The AV has been hypothesised to drive weight-related attitudes and motivate individuals to engage in unhealthy weight control practices (Hendricks, 2003; Shelley, 1997); quantitative data have supported this idea (Pugh et al., 2017;

2018). Although conceptualised as a 'voice', the AV has been differentiated from auditory verbal hallucinations (AVHs) because it is described simultaneously as a separate entity *and* a part of one's own inner speech (Higbed & Fox, 2010; Williams & Reid, 2012). Comparison of females with and without an eating disorder (ED) found that the former reported more internal dialogue about eating, weight and self-worth (Scott et al., 2014). In a study by Noordenbos and colleagues (2104), 94.5% of participants with an ED experienced a critical inner voice compared to 29.3% of healthy controls; the former also heard this voice significantly more frequently.

Researchers have drawn on aspects of dialogical self-theory (Hermans, Kempen & Van Loon, 1992) to understand the AV, suggesting it represents an internal self-critical position that attempts to dominate the more rational self (Hendricks, 2003; Williams & Reid, 2012; Williams, King & Fox, 2015). Furthermore, it has been associated with attachment theory, described as an intrapersonal, significant relationship that can affect someone's self-image (Mantailla et al., 2018a; 2018b).

The AV is said to change in nature and intensity over time (Tierney & Fox, 2010), a process reflected in findings from a meta-synthesis by Duncan et al. (2015). Initially, it appears to provide comfort and security but then becomes more sinister and powerful as AN progresses, consuming thoughts, driving behaviour and triggering negative emotions when rules and expectations are ignored (Tierney & Fox, 2010; Tierney & Fox, 2011; Williams & Reid, 2012). This change to negative appraisals of the AV, and its conceptualisation as a "voice", has been indicated as a severity marker, switching from controlled dieting to an ED (Williams & Reid, 2012; Williams et al., 2015). This is accompanied by realisations of physical emaciation, loss of control and contemplation of help seeking (Williams & Reid, 2012). However, the AV might be a barrier to accessing help because despite its negative consequences, its positive attributes may make it difficult to relinquish (Tierney & Fox, 2010; Williams et al., 2015). Attempts to encourage someone to break their bond with the AV may engender separation distress (Mantilla et al., 2018a).

It has been hypothesised that externalising anorexic thoughts as a voice may be helpful for recovery, and that therapeutic techniques facilitating acceptance of difficult internal experiences and the taking of more dominant, positive internal positions should be explored (Higbed & Fox, 2010; Williams et al., 2015; Williams & Reid, 2012). A study by Pugh and Waller (2016), using a scale developed for AVHs, found the perceived power of the AV was positively associated with ED

symptoms. However, the authors acknowledged this study potentially measured AN thoughts more generally rather than the AV specifically (Pugh & Waller, 2016). This research points to the potential importance of targeting the AV in psychological treatment, but also highlights the need for a measure developed specifically to capture the AV. To date, Dolhanty and Greenberg (2009) is the only case-study where the AV has been the focus of treatment, using emotion-focussed therapy (EFT) to decrease its harshness with moderate improvement in depression and ED symptoms.

Criticisms of the AV concept

It has been suggested the AV may be just another way of conceptualising thoughts related to AN (Fairburn, Shafran & Cooper, 1999). However, there is evidence the AV can be reliably distinguished from other inner dialogues such as self-criticism (Noordenbos, Aliakbari, & Campbell, 2014), indicating it may be a separate cognitive phenomenon (Pugh, 2016). The AV has also been criticised as being a purely social construct, emerging through research and therapeutic approaches rather than individual experience directly (Maisel, Epston & Borden, 2004). Yet qualitative research reports that individuals identify with the experience of an AV prior to contact with services (Williams et al., 2015). Finally, some have warned against the externalisation of AN as a voice, fearing it could lead to diminished responsibility over behaviour and recovery (Wright & Hacking, 2012) or give greater power to anorexic features such as the sense of being controlled (Higbed & Fox, 2010). However, qualitative descriptions unequivocally characterise the AV as both a separate entity and an integral part of the self (e.g. Williams et al., 2015), and research advocates not to 'get rid' of the AV, but rather to change the relationship between the AV and the individual (Higbed & Fox, 2010).

Objectives

There is little systematic research exploring the AV and this is difficult to progress without a valid and reliable measurement tool. To our knowledge, no existing scale adequately captures AV concepts or experiences. As such, this study aimed to develop, test and refine the Experience of an Anorexic VoicE Questionnaire (EAVE-Q). The development of this measure was informed by the findings of qualitative interview data from nine individuals with AN (Evans, 2014) alongside a review of the existing qualitative AV literature (Higbed & Fox, 2010; Tierney & Fox, 2010; Tierney & Fox, 2011;

Williams & Reid, 2012; Williams et al., 2015). This identified five recurring themes underlying the AV experience: identity and externalising (i.e. conceptualising the AV as part of the self and a separate entity); positive functions (e.g. promotion of perceived control); negative consequences (e.g. social withdrawal); power and dominance (e.g. the sense of the AV taking over the self); and the AV as a barrier to recovery (e.g. promoting secrecy). The research aims were to assess: 1) the face and content validity of the EAVE-Q through cognitive interviews; 2) the dimensions of the EAVE-Q using exploratory factor analysis (EFA); 3) the internal consistency of the EAVE-Q domains with a clinical sample; 4) construct validity by evaluating if EAVE-Q domains were significantly associated with severity markers of AN, specifically Body Mass Index (BMI), eating disorder symptoms, mood and quality of life (QoL); 5) test-retest reliability of the total scale and its subscales.

Based on the literature (e.g. Tierney & Fox, 2010) and assuming EFA would confirm the above mentioned dimensional structure of the EAVE-Q, it was predicted that any domains related to 'Positive functions of the AV' and 'Externalising the AV' would be associated with decreased distress and increased QoL, given these are linked to coping and recovery. Conversely, we predicted that domains related to 'negative consequences of the AV', 'power and dominance of the AV', and 'the AV as a barrier to recovery' would be associated with higher distress and lower QoL.

Method

Ethical approval

Ethical approval was received from the North West NHS Research Ethics Committee.

Stage 1: Scale Development

Method

Item generation

Item generation was guided by the five hypothesised AV domains, with a minimum of five marker variables per dimension to provide a scale with a stable solution (Tabachnick & Fidell, 2013). Items were developed by the authors using quotations and themes from original interview transcripts in the study by Evans (2014) and other relevant articles (Shelley, 1997; Hendricks, 2003; Higbed & Fox,

2010; Tierney & Fox, 2010; Williams & Reid, 2012; Noordenbos, Aliakbari & Campbell, 2014; Williams, King & Fox, 2015). The initial scale contained 74 items about AV experiences over the previous week and used a five point Likert scale.

Face validity, content validity and scale refinement

Cognitive interviews (Willis & Artino, 2013) were carried out with seven women with a diagnosis of AN (see Table 1) to confirm items had content and face validity and to improve the design of the measure. A standardised interview protocol was developed using existing guidelines (Willis, 2005) and included questions which probed key cognitive processes. The scale was also reviewed at a therapy group by four individuals with a diagnosis of AN receiving inpatient treatment at a specialist ED unit.

[Table 1 here]

Results

Based on consensus between participants, seven items were re-phrased and one was deleted where a consensus on clear wording could not be reached. The remaining items had high face validity. Content validity was also good, although four additional items were added to capture perceived missing aspects of the AV. The final scale had 77 items (see Table 2).

[Table 2 here]

Stage 2: Psychometric analysis

Sample

A recruitment target of 150 participants was set, which was above the absolute minimum criteria of 100 required for Factor Analysis (FA; Kline, 1994) and would allow for the reliable detection of correlations $\geq$ .22 with an alpha level of .5 and power at .80 for other planned analyses. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (Kaiser, 1974) was also calculated, with a KMO of $\geq$.5 indicating adequate sample size for FA (Hutcheson & Sofroniou, 1999). Participants aged ≥ 16 years were eligible because mid-adolescence is a key time for onset of AN (Fairburn & Harrison, 2003). Participants were required to self-report a diagnosis of AN and to obtain a global mean score

≥ 3.5 on the Eating Disorders Examination Questionnaire (EDE-Q, Fairburn & Beglin, 1994), which is within 0.5 of a standard deviation (SD) from the mean global EDE-Q score established in a previous study with a large AN help-seeking sample (Aardoom, Dingemans, Slof Op't Landt & Van Furth, 2012). BMI was not used to determine eligibility as weight restoration alone is not sufficient for recovery from AN (Fichter, Quaflieg & Hedlund, 2006; Accurso, Ciao, Fitzsimmons-Craft, Lock & Le Grange, 2014). Figure 1 shows the recruitment and screening process. The final sample included 148 participants. The overall KMO for the EAVE-Q was .85, with KMOs for individual items ranging from .69 to .92. Forty-nine participants completed the EAVE-Q at two further time points to provide an estimate of test-retest reliability.

[Figure 1 here]

Measures

Clinical history information and measures of ED symptom severity, psychological distress and QoL were used for diagnostic purposes and to assess construct validity of the EAVE-Q.

Background questionnaire

This collected relevant sociodemographic data and screened for eligibility. Information regarding predictors of AN severity were recorded, including age of onset, length of illness, BMI and total time in treatment (Schneider, Fisher, Weinerman & Lesser, 2002; Treasure & Russell, 2011).

The Eating Disorder Examination Questionnaire 6.0 (EDE-Q; Fairburn & Beglin, 1994)

A 33-item scale to assess disordered eating over a 28-day period and likely AN caseness. Higher scores indicated increased symptom severity. The EDE-Q discriminates accurately between those with and without an ED diagnosis (Aardoom et al., 2012; Berg, Peterson, Frazier & Crow, 2012). Cronbach's α (Cronbach, 1951) for EDE-Q global scores in the current sample was .81, with a mean EDE-Q score of 4.86 (SD = 0.69), which is similar to mean scores reported for AN groups in other studies where mean global EDE-Q scores >3.60 have been found (Aardoom et al., 2012; Brewin, Baggott, Dugard & Arcelus, 2014; Jennings & Phillips, 2017; Smith et al., 2017).

The Structured Clinical Interview for Diagnosis, ED Module I (SCID-DSM-5 Research Version; APA 2013)

A structured interview used to validate AN diagnosis in a sub-group of respondents following completion of the questionnaires (see 'Procedure'). A BMI of $\leq$18.5 was used to indicate 'significantly low weight' based on current guidance (National Obesity Observatory, UK, 2009). A proportion of interviews (20%) were rated independently by a second researcher to assess reliability. Inter-rater agreement for AN diagnosis was 100%.

Depression Anxiety Stress Scale (DASS-21; Lovibond & Lovibond, 1995; Appendix 11)

A 21-item scale measuring distress over the past week on three subscales (depression, anxiety and stress), with higher scores indicating more severe symptoms. Internal consistency for the DASS-21 here was high (total scale: α = .92; subscales: α = .80 to α = .92).

World Health Organisation Quality of Life Assessment (WHOQOL-BREF; Skevington, Lotfy & O-Connell, 2004; Appendix 12)

A 26-item measure assessing QoL over the previous four weeks in four domains: physical health, psychological health, social relationships and environmental. Higher scores indicate better QoL. Internal consistency here was acceptable to good, with α = .89 for the total scale, and α =.65 to .82 for individual domains.

Procedure

This was a cross-sectional, questionnaire-based study, with an optional telephone interview using the SCID-DSM-5 to confirm diagnosis at a later point. Participants were recruited through flyers distributed to public and voluntary ED services, charities and service user groups across the UK, Australia, Canada, and America. The study was also advertised on social media and online ED forums. Participants visited the study website or requested information via post (n = 2), and provided informed consent in writing or online. To confirm eligibility the background questionnaire was completed, followed by the EDE-Q. Eligible participants were directed to the main study page or sent

the study measures via post according to preference. The EAVE-Q was completed first; and the remaining measures were presented in random order. Participants provided optional consent for the SCID-DSM-5 interview or permission to contact a qualified clinician to validate AN diagnosis. All SCID interviews (n = 64) were completed by the first author and audio-recorded with participant consent. To assess test-retest reliability, participants were asked to complete the online questionnaires twice more, seven days apart. Up to three reminders were sent.

Data analysis

Exploratory factor analysis

Redundant items were identified and deleted if: a) one response category was used by >50% of respondents; b) two response categories were used by <10% of respondents; or c) the neutral response category was used by >30% of the sample (Benson & Vincent, 1980; McSharry, Bishop, Moss-Morris, Holt & Kendrick, 2015). Remaining items with inter-item correlations (IICs) >.30 and <.90 (Tabachnick & Fidell, 2013) were included for FA. Bartlett's test of sphericity was calculated, with a significant result indicating FA was likely to be meaningful. Exploratory FA was used to analyse and consolidate variables within the scale. Principal axis factoring (PAF) was used for factor extraction as items were moderately skewed (Fabrigar, Wegener, MacCallum & Strahan, 1999). Parallel analysis (PA; Horn, 1965) determined factor retention as this is the most accurate method with samples <200 (Velicer & Jackson, 1990). Only eigenvalues larger than those occurring by chance were retained. Promax rotation aided interpretation as correlations between factors were predicted. Items with no factor loadings $\geq$.40 were removed (Stevens, 2002) and items cross-loading on two or more factors $\geq$.32 were deleted to improve interpretability of the measures' dimensions (Tabachnick & Fidell, 2013). A minimum criterion of three variables per factor was specified (Costello & Osborne, 2005). Individual subscales were assessed and items with corrected item-total correlations (ITCs) <.30 were removed (Hair, Anderson, & Tatham, 1995; Tabachnick & Fidell, 2013). PAF was repeated following the removal of any variables to assess changes in the factor structure. Very high between-factor correlations were considered suggestive of overlapping dimensions better explained by a single factor and that FA was not complete. Cronbach's αs for the total scale and individual subscales were calculated.

Construct Validity

Correlations between total and subscale EAVE-Q scores and clinical outcomes (DASS-21, WHOQOL-BREF, EDE-Q and BMI) were calculated using Spearmans Rho (r_s) due to non-parametric data. The proportion of variance accounted for (R^2) was calculated for significant correlations to aid interpretation of importance.

Sensitivity analyses

This study included a proportion of participants who were below the clinical threshold for AN as eligibility was not defined by BMI or formal diagnosis. To assess how representative the sample were of a clinical population, participants with a validated diagnosis (n = 50) were compared to participants with self-report data and BMI $\leq$ 18.5 (n = 54). Differences between groups for continuous variables (age, BMI, DASS-21 scores, WHOQOL-BREF scores, EDE-Q scores) were compared using the independent t-test. Differences in categorical variables (gender, current treatment and length of treatment) were explored using Chi Square analyses. The ability of the EAVE-Q to discriminate between those above and below clinical threshold for AN was assessed within the full sample using the t-test to compare EAVE-Q scores when participants were separated into a 'clinical' group (validated diagnosis of AN or BMI $\leq$ 18.5; n = 104) and a 'sub-threshold' group (BMI >18.5; n = 44).

Test-retest reliability

The intraclass correlation (ICC) between sub-scale scores and total EAVE-Q scores collected at Time 1 and Time 2 was calculated. A two-way mixed effects model with single measurement and absolute agreement was selected (Koo & Li, 2016). ICCs <0.5 were considered poor, 0.5 - 0.75 moderate, 0.76 - 0.9 good and >0.90 considered to indicate excellent reliability (Portney & Watkins, 2000).

Results

Sample characteristics

Sample characteristics are presented in Table 3.

[Table 3 here]

Exploratory Factor analysis

After removing 44 redundant items, all remaining items had IICs >.30 and <.90 and were included for FA. Bartlett's test of sphericity was highly significant ($p < .0001$), indicating FA was appropriate. Items retained for FA used the entire response scale, indicating good scale to sample targeting. PA indicated seven initial factors occurring above chance (see Figure 2). Items with no factor loadings $\geq$.40 were removed (n = 7). Two of the initial seven factors had only two items loading $\geq$.40 and were deleted (n = 4 items). Four items cross-loading $\geq$.32 were also deleted. In total eight PAF analyses were required to find the most interpretable factor structure with the fewest items and adequate internal consistency. PAF analyses were run with seven, six, five and four factors selected for extraction with the remaining 18 items. The five-factor solution provided the best fit for the data (see Table 4), with 54.83% of the total variance explained and average communalities of .55.

[Table 4 here]

[Figure 2 here]

Interpreting factors

Although the five-factor model was consistent with the proposed theoretical model of the EAVE-Q, the factors extracted were not identical to those predicted. Factors 1 and 2 encompassed hypothesised benefits of the AV, but were clearly distinguishable dimensions. Factor 1, labelled 'Benefits of adhering to the AV', related to the positive consequences of obeying the AV's demands and was linked to perceived control and positive emotions. Factor 2, labelled 'The compassionate AV', represented a more supportive function, linked to feelings of being comforted and understood by the AV. As predicted, Factor 3 encompassed items depicting the AV as a barrier to recovery. However, items loading onto this factor were specifically linked to mistrust and distancing the self from others, and so was re-labelled 'Turning away from others'. Factor 4 was as hypothesised and retained the label 'Externalising the AV', capturing the extent to which individuals identified the AV as part of, or external to, themselves. Similarly, Factor 5 had items loading onto it which were proposed to

underpin a powerful and dominant domain and so retained the label 'Dominated by the AV'. The hypothesised 'Negative consequences of engaging with the AV' was not found, with these aspects of the AV potentially subsumed within other factors.

Reliability

Cronbach's α for the full scale was α = .83. All subscales had good internal consistency (see Table 5), with ITCs >.30 and α ranging from α = .70 to α = .85.

[Table 5 here]

Construct validity

Moderate correlations between factors were predicted and suggested EAVE-Q domains measured a shared, higher order construct (the AV). No two factors were so highly correlated that they were likely to have been better explained by a single factor (see Table 6). In general, factors had small to moderate correlations, as would be expected in a multidimensional scale. 'Benefits of adhering to the AV' and 'The compassionate AV' domains were largely correlated, which is consistent with the hypothesis that these are both positive functions of the AV. However, items did not cross-load on to the other dimension >.32 and so were retained as separate factors. 'The compassionate AV' dimension correlated moderately with 'Turning away from others', indicating the experience of a supportive AV was associated with people withdrawing from external support. 'Externalising the AV' correlated to some extent with 'The compassionate AV' and the 'Dominated by the AV' domains. 'Externalising the AV' and 'Benefits of adherence' were uncorrelated, suggesting the extent to which individuals see the AV as part of themselves or as a separate entity does not relate to positive appraisals of the AV experience.

[Table 6 here]

Associations between EAVE-Q domains and clinical outcomes

Significant correlations between EAVE-Q domains and other clinical outcomes ranged from small to moderate (see Table 7). The 'Dominated by the AV', 'Turning away from others' and 'The compassionate AV' subscales were all moderately and positively correlated with ED symptoms. 'Benefits of adherence' was also positively correlated with EDE-Q scores, although to a lesser extent. No correlations accounted for more than 50% variance in EDE-Q scores. All EAVE-Q domains were significantly positively associated with the EDE-Q Restraint, Shape Concern and Weight Concern subscales. Eating concern was also significantly associated with higher scores on the 'Turning away from others' and 'Dominated by the AV' domains. Small but significant correlations were found between EDE-Q vomiting and compulsive exercise behaviours and the 'Dominated by the AV' subscale.

Age of onset was negatively correlated with the 'Turning away from others' subscale, whilst 'The compassionate AV' subscale was negatively associated with years since first onset. There was no difference in EAVE-Q scores on any domain between individuals identified as AN-R and AN-BP using the SCID (see Table 8).

'Turning away from others' was positively and moderately associated with DASS-21 scores. All domains on the DASS-21 had small, but significant, positive correlations with scores on the 'Dominated by the AV' dimension, with correlations approaching a moderate association for the DASS-21 stress subscale. 'The compassionate AV' domain correlated positively with anxiety scores on the DASS-21 only, whilst 'Benefits of adherence' did not correlate well with any subscales on the DASS-21.

'Turning away from others' was significantly and negatively associated with all domains of the WHOQOL-BREF, being most associated with scores on the psychological domain which includes questions about self-acceptance and the meaning and enjoyment of life. All WHOQOL-BREF domains correlated negatively with 'Dominated by the AV' scores, most notably the psychological domain and the social domain, which records satisfaction with relationships and support. In contrast, the 'Benefits of adherence' scale had small, positive associations with WHOQOL-BREF physical, psychological and environmental domains. Only the environmental domain achieved significance, which includes questions about access to health services, financial security and opportunities for leisure activities. 'The compassionate AV' had small negative correlations with psychological and social domains on the

WHOQOL-BREF, but these were not approaching significance. BMI was not significantly associated with any of the EAVE-Q domains. The 'Externalising the AV' scale was not correlated with any clinical outcomes.

[Table 7 here]

[Table 8 here]

Sensitivity analyses

Sixty-four participants completed the SCID and five more had diagnosis verified by their health professional. Of those, 50 participants met full criteria for AN; 36 for restricting type and 14 for binge-purge type. Fourteen people were sub-threshold for AN on the SCID due to having BMI >18.5 or failing to meet full criteria for fear of weight gain or disturbance of weight or shape. Participants with a validated diagnosis of AN (n = 50, mean age = 32.14 years) were significantly older than those who were underweight with self-report data only (n = 54, mean age = 25.35 years; t(102) =3.14, p = .002; MD = 6.79, 95% CI = 2.49 - 11.08). However, groups did not differ significantly on any other outcomes, suggesting similar severity of AN and psychological distress. This indicates that participants with self-report data and BMI $\leq$18.5 were generally representative of a clinical AN sample. There was no significant difference in EAVE-Q scores for the 'clinical group' (participants with a validated diagnosis or BMI $\leq$18.5; n = 104) and those who were 'sub-threshold' for AN (BMI >18.5; n = 44) (see Table 9). Therefore, whilst linked to ED symptoms and psychological distress, endorsement of AV experiences was not associated with lower weight. Finally, one outlier (>3 times the interquartile range) was identified based on mean EAVE-Q scores. All analyses were completed with and without this person's data; no difference in the factor structure of the EAVE-Q was found, which appeared robust within this sample.

[Table 9 here]

Test-retest reliability

Average time between Time 1 and Time 2 EAVE-Q completion was 29 days (see Table 10). ICCs for the total scale score and individual EAVE-Q domains were all within the 'moderate' range.

[Table 10 here]

Discussion

This paper describes the development of the first scale measuring the AV. Exploratory FA resulted in 18 items organised into five domains to form the EAVE-Q. As predicted, domains pertaining to 'Externalising the AV' and 'Dominated by the AV' were identified, capturing the duality and powerlessness associated with experiences of the AV reported in the literature (e.g. Tierney & Fox, 2010). The hypothesised domain of 'The AV as a barrier to recovery' was not found, but instead a more specific dimension of 'Turning away from others' was elicited and underpinned by beliefs about trust and self-worth in relationships. The hypothesised dimension of the AV having a positive function was identified but was better conceptualised as two separate domains. 'The compassionate AV' encapsulated a supportive function of the AV, whilst 'Benefits of adhering to the AV' encompassed positive consequences of engagement. The hypothesised negative consequences of the AV domain had unexpected nuances and was differentiated as negative internal consequences in the 'Dominated by the AV' domain, and negative social consequences in the 'Turning away from others' dimension. Internal consistency for the EAVE-Q was high for this sample, who predominantly met clinical threshold for AN, indicating that the scale was reliably measuring the AV construct.

Test-retest reliability was moderate and construct validity was good. Except for the 'Externalising the AV' subscale, all dimensions were significantly and positively correlated with EDE-Q scores, suggesting that the AV is an important clinical feature of AN. Other researchers have highlighted that feeling overpowered and controlled by the AV, and submitting to this, is associated with greater eating pathology (Mantilla et al., 2018b; Pugh & Waller, 2017). Likewise, positive beliefs about the AV have been associated with disordered eating attitudes (Pugh et al., 2018), in line with our findings on the domain 'Benefits of adhering to the AV'.

No correlations accounted for more than 50% variance in EDE-Q scores, indicating that the AV is not just another way of conceptualising cognitive and behavioural symptoms of AN (Hinken,

Tracey & Enz, 1997). There were significant correlations between EAVE-Q scores and measures of eating restraint, vomiting, compulsive exercise, shape concern and weight concern, supporting assertions that the AV is linked to weight-related attitudes and weight management behaviour (Higbed & Fox, 2010; Tierney & Fox, 2010) although causality cannot be assumed due to the cross-sectional nature of this study.

As predicted, there were significant positive correlations between EAVE-Q domains, psychological distress and AN severity markers. The 'Turning away from others' and 'Dominated by the AV' domains were moderately associated with depression, anxiety and stress, and reduced QoL. The largest effects were found for associations with psychological and social WHOQOL-BREF domains. This is logical, as the social domain measures satisfaction with relationships, which would not be expected for people endorsing the 'Turning away from others' subscale. The psychological domain assesses self-acceptance and satisfaction with weight and shape, which would also not be predicted for people endorsing the 'Dominated by the AV' scale, as perceptions of being powerless to resist the AV have been associated with feelings of failure, guilt and shame (Tierney & Fox, 2011). 'Benefits of adherence' had small positive associations with three of the four WHOQOL-BREF subscales, although only associations with the environmental subscale reached significance. This may be due to this WHOQOL subscale including questions about satisfaction with access to health services, as 66% of the sample was receiving current treatment. Age of onset was moderately and negatively correlated with the 'Turning away from others' subscale, with younger age of onset associated with individuals endorsing their AV prevented them from trusting others and feeling worthy of help. 'The compassionate AV' subscale was also negatively associated with years since first onset, suggesting that those who had lived with AN symptoms the longest were less likely to endorse the compassionate aspects of their AV.

There were no significant associations between 'Benefits of adherence' and DASS-21 scores, or 'The compassionate AV' and QoL scores. This is surprising, as the literature would suggest that the positive functions of the AV are the avoidance of painful emotions and the protection of self-esteem (e.g. Williams & Reid, 2012), and so negative relationships between these domains and distress, and positive associations with QoL were expected. There was also a small positive correlation between the DASS-21 anxiety subscale and 'The compassionate AV' subscale, and a small negative correlation

between 'The compassionate AV' domain and QoL, which was counter-intuitive. The size of these associations indicates future research with a larger sample is required to assess if they are robust. If so, this could be reflective of the tension between the AV's positive and negative aspects, and the potential insecure attachments many individuals have with their EDs (Mantilla et al., 2018b). It is possible that as anxiety increases, endorsement of the AV as supportive and comforting also increases, as people turn to a trusted and relied upon coping strategy (Tierney & Fox, 2010).

The literature suggests that externalising the AV could be a key part of recovery (Higbed & Fox, 2010); therefore, positive associations with QoL and negative associations with psychological distress and ED symptoms were predicted. No such associations were found, raising questions about the clinical utility of this domain. Limited sample size and the non-parametric distribution of items within the scale may mean some items lacked variability and more subtle associations were missed. Whether associated with clinical outcomes or not, perhaps a more relevant question is the utility of this domain in understanding the experience of the AV. As 'Externalising the AV' was a robust dimension in FA and in the qualitative literature underpinning this study, it was retained for further analysis in future research.

Limitations

The non-parametric distribution of individual EAVE-Q items means that results of FA are restricted to the study sample (Field, 2005). The sample was predominantly White British (70%) and female (97%), meaning results may not be generalisable to individuals with AN from other cultures or males. The EAVE-Q did not distinguish between clinical and sub-threshold participants based on BMI, which is perhaps not surprising given that only participants with clinically relevant scores on the EDE-Q were included, and that psychological recovery does not co-occur with weight restoration alone (Fichter et al., 2006; Le Grange et al., 2013). Others have also noted no association between characteristics of the AV and BMI (Pugh et al., 2018). However, the accuracy of self-reported weight is a potential confound, as research regarding accuracy of weight reporting is contentious for clinical and recovered AN samples (McCabe, McFarlane, Polivy & Olmsted, 1999; Wolfe, Kelly-Weeder, Malcolm & McKenery, 2013). Participants also self-reported diagnosis of AN, although attempts were made to check this via

SCID interview or confirmation with health professionals, and sensitivity analyses suggested no significant clinical differences between those with or without a validated diagnosis of AN.

Clinical Implications

Other authors have highlighted how severity of eating pathology can be associated with appraisal of the AV as strong and powerful (Pugh & Waller, 2017; Pugh et al., 2018). Our study provides further evidence for the clinical importance of the AV in AN. Higher EAVE-Q scores were associated with negative ED attitudes, increased psychological distress, and decreased QoL. The EAVE-Q could be used to introduce discussions about the AV within clinical settings, promoting shared understandings and formulations between clients and professionals. The EAVE-Q could also be used to assess the ability of existing interventions to target the AV.

Implications for Future Research

This study provides a refined scale ready for validation in other AN samples. Further research is required to determine: a) if the 'Externalising the AV' domain is a key facet of the AV, and to clarify patterns of associations with clinical outcomes and their meaning, and b) to what extent the AV, as assessed by the EAVE-Q, mediates clinical outcomes, such as ED symptoms, QoL and use of services. Considering the lack of recent studies exploring novel treatment strategies in AN (Lipsman, Woodside & Lozano, 2014), the development and assessment of psychological approaches targeting the AV is a worthwhile pursuit. A potential starting point is EFT (Greenberg & Johnson, 1988), as there is evidence that EFT techniques addressing the harshness of the AV can reduce depression and ED symptoms (Dolhanty & Greenberg, 2009). The EAVE-Q could be used to measure if reductions in AV endorsement mediate these outcomes. Narrative therapy has also been proposed as a means of enabling people to explore more adaptive attachment relationships outside of their ED (Mantilla et al., 2018a); again, the EAVE-Q could be used to examine the validity of this claim. Furthermore, compassion-focused approaches (Goss & Allan, 2014) could enable individuals to be more kind and forgiving to their own perceived failings, counteracting negative internal encounters shaped by how they respond to the AV.

Conclusions

When considering new directions for future research and treatment in AN, it is important to consider the priorities and perspectives of those experiencing this ED. The EAVE-Q is the first stage in the development of a reliable and valid measure of the AV based on experiences of individuals with a diagnosis of AN. The EAVE-Q provides a tool to aid the continuing development of patient-centred research and effective interventions in AN.

References

Aardoom, J. J., Dingemans, A. E., Slof Op't Landt, M. CV. & Van Furth, E. F. (2012). Norms and discriminative validity of the Eating Disorder Examination Questionnaire (EDE-Q). *Eating Behaviors, 13* (4), 305 – 309. doi.org/10.1016/j.eatbeh.2012.09.002

Accurso, E. C., Ciao, A. C., Fitzsimmons-Craft, E. E., Lock, J.D., & Le Grange, D. (2014). Is weight gain really a catalyst for broader recovery? The impact of weight gain on psychological symptoms in the treatment of adolescent anorexia nervosa. *Behaviour Research and Therapy, 56,* 1 – 6. doi.org/10.1016/j.brat.2014.02.006

Alwin, D. E. (1992). Information transmission in the survey interview: Number of response categories and the reliable attitude measurement. In P. V. Marsden (Ed.), *Sociological Methodology,* (pp. 83 – 118). Washington DC: John Wiley & Sons, Inc.

Alwin, D.F. (1997). Feeling thermometers versus 7 -point scales. *Sociological Methods and Research, 25* (3), 318 - 351.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Publishing.

Ashford, S., Turner-Stokes, L., Siegert, R., & Slade, M. (2013). Initial psychometric evaluation of the Arm Activity Measure (ArmA): A measure of activity in the hemiparetic arm. *Clinical Rehabilitation*, *27*, 728–740. doi.org/10.1177/0269215512474942

Baiano, M., Salvo, P., Righetti, P., Cereser, L., Baldissera, E., Camponogara, I. & Balestrieri, M. (2014). Exploring health-related quality of life in eating disorders by a cross-sectional study and a comprehensive review. *BMC Psychiatry*, *14* (165), 1 – 12. doi.org/10.1186/1471-244X-14-165

Benson, P.L. & Vincent, S. (1980) Development and validation of the sexist attitudes toward women scale (SATWS). *Psychology of Women Quarterly, 5,* 276 – 291. doi.org/10.1111/j.1471-6402.1980.tb00962.x

Berg, K. C., Peterson, C. B., Frazier, P. F. & Crow, S. J. (2012). Psychometric evaluation of the eating disorders examination and eating disorders examination questionnaire: A systematic review of the literature. *International Journal of Eating Disorders, 45* (3), 428 – 438. doi.org/10.1002/eat.20931

Brewin, N., Baggott, J., Dugard P. & Arcelus, J. (2014). Clinical Normative Data for Eating Disorder Examination Questionnaire and Eating Disorder Inventory for DSM-5 Feeding and Eating Disorder Classifications: A Retrospective Study of Patients Formerly Diagnosed Via DSM-IV. *European Eating Disorders Review,* 22(4):299–305. doi.org/10.1002/erv.2301.

Costello, A. B. & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation, 10* (7), 1 – 9. Retrieved from http://pareonline.net/getvn.asp?v=10&n=7

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16* (3), 297 – 334.

Duncan, T.K., Sebar, B. & Lee, J. (2014). Reclamation of power and self: a meta-synthesis exploring the process of recovery from anorexia nervosa*. Advances in Eating Disorders: Theory, Research and Practice*, 3, 177–190. doi.org/10.1080/21662630.2014.978804

Evans, G. F. (2014). *The anorexic voice: Characterisation and measurement of voice hearing in anorexia nervosa* (unpublished masters thesis). University of Manchester, UK.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C. & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4* (3), 272 – 299. Retrieved from http://ww.w.statpower.net/Content/312/Handout/Fabrigar1999.pdf

Fairburn, C. G. & Beglin, S. J. (1994). Assessment of eating disorder psychopathology: interview or self-report questionnaire? *International Journal of Eating Disorders*, *16*, 363 – 370. doi.org/10.1002/1098-108X

Fairburn, C. G. & Harrison, P. J. (2003). Eating disorders. *Lancet*, *361,* 407 -416. doi.org/10.1016/S0140-6736(03)12378-1

Fairburn, C. G., Shafran, R. & Cooper, Z. (1999). A cognitive behavioural theory of anorexia nervosa. *Behaviour Research & Therapy, 37,* 1-13. doi.org/10.1016/S0005-7967(98)00102-8

Fan, X. & Sivo, S. A. (2007). Sensitivity of fit indices to Model Misspecification and Model Types. *Multivariate Behavioral Research, 42 (3),* 509 – 529. doi.org/10.1080/00273170701382864

Fichter, M. M., Quaflieg, N. & Hedlund, S. (2006). Twelve-year course and outcome predictors of anorexia nervosa. *International Journal of Eating Disorders, 39*, 87 – 100.

Field, A. (2005). Exploratory Factor Analysis. In A. Field (Ed.), *Discovering Statistics Using SPSS* (pp. 619 – 679). London: Sage Publications Ltd.

Goss, K. & Allan, S. (2014). The development and application of compassion-focused therapy for eating disorders (CFT-E). *British Journal of Clinical Psychology, 53,* 62-77. 10.1111/bjc.12039

Greenberg, L. S. & Johnson, S. M. (1988). *Emotionally focused therapy for couples*. New York: Guilford Press.

Hair, J., Anderson, R. E. & Tatham, R. L. (1995). *Multivariate data analysis*. New Jersey: Prentice-Hall Inc.

Hendricks, J. (2003). *Slim to none: A journey through the wasteland of anorexia treatment.* New York: McGraw-Hill Education.

Hermans, H. J. M., Kempen, H. J. G. & Van Loon, R. J. P. (1992). The dialogical self: Beyond individualism and rationalism. *American Psychologist, 47,* 23-33.

Higbed, L. & Fox, J. R. (2010). Illness perceptions in anorexia nervosa: a qualitative investigation. *British Journal of Clinical Psychology, 49* (3), 307-325. doi.org/10.1348/014466509X454598

Hinken, T. R., Tracey, J. B. & Enz, C. A. (1997). Scale construction: Developing reliable and valid measurement instruments. *Journal of Hospitality and Tourism Research, 21* (1), 100 – 120. doi.org/10.1177/109634809702100108

Horn, J. L. (1965). A rationale and technique for estimating the number of factors in factor analysis. *Psychometrika, 30*, 179-185. doi.org/10.1007/BF02289447

Hutcheson, G. D., & Sofroniou, N. (1999). *The Multivariate Social Scientist: an introduction to generalized linear models.* New York: Sage Publications.

Jennings, K. M. & Phillips, K. E. (2017). Eating Disorder Examination–Questionnaire (EDE–Q): Norms for Clinical Sample of Female Adolescents with Anorexia Nervosa. *Archives of Psychiatric Nursing,* 31(6): 578–581*. doi.org/10.1016/j.apnu.2017.08.002*

Kaiser, H. F. (1974). An index of factorial simplicity*. Psychometrika*, 39, 31-36. doi.org/10.1007/BF02291575

Kline, P. (1994). *An Easy Guide to Factor Analysis*. New York: Routledge.

Kline, P. (1999). *The Handbook of Psychological Testing*. London: Routledge.

Le Grange, D., Crosby, R. D., Engel, S. G., Cao, L., Ndungu, A., Crow, S. J., Peterson, C. B., Mitchell, J. E. & Wonderlich, S. A. (2013). DSM-IV-Defined Anorexia Nervosa Versus Subthreshold Anorexia Nervosa (EDNOS-AN). *European Eating Disorders Review, 21* (1), 1 – 7. doi.org/10.1002/erv.2192

Lovibond, P. F. & Lovibond, S. H. (1995). The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behavior Research and Therapy, 33* (3), 335 – 343. doi.org/10.1016/0005-7967(94)00075-U

Maisel, R., Epston, D. & Borden, A. (2004). *Biting the hand that starves you: Inspiring resistance to anorexia/ bulimia*. New York: Norton and Company.

Mantilla, E.F., Clinton, D. & Birgegard, A. (2018a). The unsafe haven: Eating disorders as attachment relationships. *Psychology and Psychotherapy: Theory, Research and Practice. 92(3):379-393. doi: 10.1111/papt.12184*

Mantilla, E.F., Clinton, D. & Birgegard, A. (2018b). Insidious: The relationship patients have with their eating disorders and its impact on symptoms, duration of illness, and self-image. *Psychology and Psychotherapy: Theory, Research and Practice*, 91, 302–316. DOI: 10.1111/papt.12161

McCabe, R. E., McFarlane, T., Polivy, J. & Olmsted, M. P. (2001). Eating disorders, dieting, and the accuracy of self-reported weight. *International Journal of Eating Disorders, 29* (1), 59 – 64. doi.org/10.1002/1098-108X

McSharry, J., Bishop, F. L., Moss-Morris, R., Holt, R. I. & Kendrick, T. (2015). A new measure of multimorbid illness and treatment representations: the example of diabetes and depression. *Journal of Affective Disorders, 174,* 192 – 200. doi.org/10.1016/j.jad.2014.11.050

Miller, G (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review, 63,* 81 – 97.

National Obesity Observatory (2009): http://www.noo.org.uk/uploads/doc789_40_noo_BMI.pdf. Accessed on 1st June 2015.

National Institute for Clinical Excellence. (2004). *Eating disorders in over 8s: Management.* NICE guideline: CG9. https://www.nice.org.uk/guidance/cg9

Noordenbos, G., Aliakbari, N., & Campbell, R. (2014). The relationship among critical inner voices, low self-esteem, and self-criticism in eating disorders. *Eating Disorders: The Journal of Treatment and Prevention, 22*, 337–351. · doi.org/10.1080/10640266.2014.898983

Portney L, G. & Watkins M. P. (2000). *Foundations of clinical research: applications to practice*. Prentice Hall; New Jersey.

Preston, C. C. & Colman, A. M. (2000). Optimal number of response categories in rating scales – reliability, validity, discriminating power and respondent preferences. *Acta psychologica*, 104, 1 – 15. doi.org/10.1016/S0001-6918(99)00050-5

Pugh, M., Waller, G. & Espositoc, M. (2018). Childhood trauma, dissociation, and the internal eating disorder 'voice'. *Child Abuse and Neglect, 86,* 197-205. 10.1016/j.chiabu.2018.10.005

Pugh, M. (2016). The internal 'anorexic voice': A feature or fallacy of eating disorders? *Advances in Eating Disorders: Theory, Research & Practice*, *4* (1), 73 – 83. doi.org/10.1080/21662630.2015.1116017

Pugh, M. & Waller, G. (2017). Understanding the 'Anorexic Voice' in Anorexia Nervosa. *Clinical Psychology and Psycho*therapy, 24, 670–676. doi.org/10.1002/cpp.2034

Pugh, M. & Waller, G. (2016). The anorexic voice and severity of eating pathology in anorexia nervosa. *International Journal of Eating Disorders, 49* (6), 622-625. doi.org/10.1002/eat.22499

Schneider, M., Fisher, M., Weinerman, S. & Lesser, M. (2002). Correlates of low bone density in females with anorexia nervosa. *International Journal of Adolescent Health and Medicine, 14* (4), 297 – 306. doi.org/10.1515/IJAMH.2002.14.4.297

Scott, N., Hanstock, T.L. & Thornton C. (2014). Dysfunctional self-talk associated with eating disorder severity and symptomatology. *Journal of Eating Disorders*, *2*, 14.doi.org.10.1186/2050-2974-2-14

Shelley, R. (1997). *Anorexics on anorexia.* London: Jessica Kingsley Publishers Ltd.

Skevington, S. M., Lotfy, M., & O-Connell, K. A. (2004). The World Health Organisation's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. *Quality of Life Research, 13* (2), 299 – 310. doi.org/10.1023/B:QURE.0000018486.91360.00

Smith, K. E., Mason, T. B., Murray, S. B, Griffiths, S., Leonard, R. C., Wetterneck, C. T., Smith, B. E. R., Farrell, N. R., Riemann, B. C. & Lavender, J.M (2017). Male clinical norms and sex differences on the eating disorder inventory (EDI) and eating disorder examination questionnaire (EDE-Q). *The International Journal of Eating Disorders*, 50 (7):769–775. Doi.org.10.1002/eat.22716

Stevens J. (2002). *Applied Multivariate Statistics for the Social Sciences*. Mahwah, New Jersey: Erlbaum.

Tabachnick, B. G. & Fidell, L. S. (2013). *Using Multivariate Statistics.* (6th ed.). Boston: Pearson.

Tierney, S. & Fox, J. R. (2010). Living with the anorexic voice: a thematic analysis. *Psychology and Psychotherapy, 83* (3), 243-254. doi.org/10.1348/147608309X480172

Tierney, S. & Fox, J. R. (2011). Trapped in a toxic relationship: comparing the views of women living with anorexia nervosa to those experiencing domestic violence. *Journal of Gender Studies, 20* (1), 31-41. doi.org/10.1080/09589236.2011.542018

Treasure, J. & Russell, G. (2011). The case for early intervention in anorexia nervosa: theoretical exploration of maintaining factors. *The British Journal of Psychiatry, 199* (1), 5 – 7. doi.org/10.1192/bjp.bp.110.087585

Velicer, W. F. & Jackson, D. N. (1990a). Component analysis versus common factor analysis: Some issues in selecting an appropriate procedure. *Multivariate Behavioral Research, 25*, 1 - 28. doi.org/10.1207/s15327906mbr2501_1

Williams, K., King, J. & Fox, J. R. (2015). Sense of self and anorexia nervosa: A grounded theory. *Psychology & Psychotherapy, 89* (2), 211-228. doi.org/10.1111/papt.12068

Williams, S. & Reid, M. (2012). "It's like there are two people in my head": a phenomenological exploration of anorexia nervosa and its relationship to the self. *Psychology & Health, 27* (7), 798-815. doi.org/10.1080/08870446.2011.595488

Willis, G. B. (2005). *Cognitive Interviewing: A tool for improving questionnaire design.* (1st ed.). Thousand Oaks, CA: Sage Publications, Inc.

Willis, G. B. & Artino, A. R. (2013). What do our respondents think we're asking? Using cognitive interviewing to improve medical education surveys. *Journal of Graduate Medicine and Education, 5* (3), 353 – 356. doi.org/10.4300/JGME-D-13-00154.1

Wolfe, B. E., Kelly-Weeder, S., Malcom, A. W. & McKenery, M. (2013). Accuracy of self-reported body weight and height in remitted anorexia nervosa. *Journal of the American Psychiatric Nurses Association, 19* (2), 66 - 70. doi.org/10.1177/1078390313481062

Wright, K. M. & Hacking, S. (2012). An angel on my shoulder: A study of the relationships between women with anorexia and healthcare professionals. *Journal of Psychiatric and Mental Health Nursing, 19,* 107-115. doi.org/10.1111/j.1365-2850.2011.01760.x

Table 1. Stage 1 cognitive interview sample characteristics

Variable	**n = 7**
Age (years)	
Mean	26
Range	18 – 54
Ethnicity:	
White British	7

Current treatment for AN	7
BMI < 18.5	7
Place of interview	
Participant's home	2
Community location	3
Inpatient ward	2

Table 2. EAVE-Q Items prior to EFA (n = 77)

Identity & Externalising (n = 7)	**Negative consequences of engaging with the AV (n = 13)**	**The AV as a barrier to recovery (n = 14)**	**Power & dominance of the AV (n = 24)**	**Positive functions of the AV (n = 19)**
* I experience my AV as thoughts that are not my own	* My AV misleads me	* My AV tells me that it's the only thing I should rely on	* My AV is a voice in my head which influences how I think and feel	* My AV is supportive
* My AV is like hearing someone else's thoughts and feelings	* My AV makes me feel weak	* I want to get rid of my AV	* My AV is a feeling inside which influences what I can and can't do	* My AV is comforting
* My AV is part of who I am	* My AV is a negative part of my life	* My AV makes the things I do to lose weight seem OK	* My AV is always there	* My AV is a friend to me
* I see my AV as separate from my own identity	* My AV is harmful to me	* I can control my AV	* My AV tells me I should be punished	* My AV understands me when other people don't
* My AV is always the same gender	* My AV makes me act like someone I don't want to be	* My AV makes me think other people just want me to get fat	* My AV is so loud it's hard to hear any other thoughts	* My AV helps me to cope when things are difficult
* My AV reminds me of someone I know	* My AV makes me feel isolated	* My AV tells me not to trust other people	* My AV gets louder when I'm feeling stressed or down	* My AV is a positive part of my life
* I see my AV as something that shouldn't be there	* I feel angry with my AV	* My AV motivates me to want to lose weight	* I believe everything my AV says	* My AV makes me feel in control
	* My AV makes me feel angry or frustrated with myself	* My AV makes me secretive and hide things from other people	* My AV becomes louder when I ignore it	* My AV gives me a positive sense of routine and order in my life
	* My AV makes me feel angry or frustrated with other people	* I can ignore my AV if I really want to	* My AV gets louder when I eat something it says I shouldn't	* Doing what my AV says makes me feel satisfied
	* My AV makes me think that life isn't worth living	* My AV makes me think I don't deserve other people's help	* My AV bullies me	* Doing what my AV says makes me feel happy
	* My AV distresses me	* My AV makes me believe I don't deserve food	* My AV controls me	* My AV gives me a sense of purpose

Identity & Externalising (n = 7)	**Negative consequences of engaging with the AV (n = 13)**	**The AV as a barrier to recovery (n = 14)**	**Power & dominance of the AV (n = 24)**	**Positive functions of the AV (n = 19)**
	* My AV makes it hard for me to maintain relationships with others	* My AV advises me to get help	* My AV criticises me	* My AV gives me a sense of achievement
	* When I hear my AV negative images come into my mind	* I need to get rid of my AV to feel like I have completely recovered	* My AV is powerful	* My AV makes me a better person
		* Treatment should help me manage my AV	* My AV takes over me	* My AV makes me feel less alone
			* My AV is strong	* My AV makes me feel safe
			* My AV gets stronger when I gain weight	* My AV makes me feel confident
			* My AV gets stronger when I lose weight	* I feel lonely when my AV is not there
			* My AV makes me feel anxious or panicky when I go against it	* When I hear my AV comforting images come into my mind
			* My AV makes me feel guilty when I go against it	* My AV helps me block out painful thoughts and feelings
			* My AV gives me rules and regulations I have to follow	
			* My AV makes me feel ashamed when I go against it	
			* It doesn't matter what I do, my AV always wins	
			* Not doing what my AV tells me makes me feel exhausted	
			* I feel like I'm battling against my AV	

Table 3. Stage 2 sample characteristics (n = 148)

Variable		n / value
Age	Mean	27.74 years
	Range	16 – 63
Gender ratio	M/F	4 / 144
Ethnicity	White	127
	Asian	2
	Chinese	2
	Hispanic	2
	Black Caribbean	1
	Other	2
	Missing	12
Country of origin	Europe (UK)	113 (105)
	USA	26
	Canada	4
	Australia	3
	Other	2
Education	GCSE or equivalent	27
	A-level or equivalent	34
	Degree level or equivalent	67
	Missing	20
Age at first AN symptoms	Mean	15.09 years
	Range	5 – 39 years
Age at diagnosis	Mean	19.95 years
	Range	11 – 49
	Missing	n = 1
Current treatment for AN	Yes	100
	No	48

Ever treated for AN	Yes	141
	No	7
Length of AN treatment	Never treated	7
	< 1 year	20
	1 – 2 years	38
	3 – 4 years	25
	5 – 6 years	17
	> 6 years	39
	Missing	2
Current BMI	Mean (SD)	17.61 (2.44)
	Range	13.17 – 25.53
EDE-Q: Mean (SD)	Global score	4.86 (0.69)
DASS-21: Mean (SD)	Total score	36.23 (12.85)
	Depression	13.46 (5.70)
	Anxiety	9.42 (4.98)
	Stress	13.35 (4.57)
WHOQOL-BREF: Mean (SD)	Physical	51.84 (16.99)
	Psychological	25.39 (16.59)
	Social relationships	39.04 (22.07)
	Environment	61.46 (17.46)
	Q1: Quality of life	2.93 (1.02)
	Q2: Health	2.29 (0.96)

Table 4. Stage 2 promax pattern matrix for the EAVE-Q five factor solution with 18 items (n = 148)

	Factor				
Item	**1**	**2**	**3**	**4**	**5**
My AV makes me feel in control	**.837**	-.063	-.082	.096	-.106
My AV gives me a positive sense of routine and order in my life	**.754**	-.011	.033	.021	-.051
Doing what AV says makes me feel happy	**.630**	.029	.055	-.116	.044
Doing what AV says makes me feel satisfied	**.626**	.057	-.010	-.047	.135
My AV makes me feel confident	**.499**	.111	-.032	-.024	-.042
My AV is supportive	-.091	**.826**	-.005	.058	-.123
My AV is comforting	.094	**.788**	-.137	.020	.047
My AV is a friend to me	.098	**.759**	.079	-.004	-.035
My AV understands me when other people don't	.114	**.548**	.140	-.058	.104
My AV makes me think other people just want me to get fat	.094	-.127	**.899**	-.007	-.137
My AV tells me not to trust other people	-.053	.071	**.861**	.017	.033
My AV makes me think I don't deserve other people's help	-.123	.098	**.455**	.021	.137
I experience my AV as thoughts that are not my own	.075	-.073	.011	**.922**	.023
I see my AV as separate from my own identity	-.134	.113	-.071	**.672**	-.099
My AV is like hearing someone else's thoughts and feelings	.024	.025	.121	**.576**	.132
It doesn't matter what I do my AV always wins	.047	-.062	-.003	-.029	**.795**
I can ignore my AV if I really want to (reversed)	-.167	-.013	.002	-.022	**.639**
My AV controls me	.102	-.001	-.031	.089	**.606**

NB Extraction method: Principal axis factoring; Rotation: Promax with Kaiser normalisation. Items in bold type and underlined are items loading onto each factor ≥.40

Table 5. Cronbach's alpha reliability coefficients of the EAVE-Q scales derived from FA

Dimension	Number of items	α	ITC range
1. Benefits of adhering to the AV	5	0.81	0.48 – 0.66
2. The compassionate AV	4	0.85	0.62 – 0.77
3. Turning away from others	3	0.78	0.47 – 0.73
4. Externalising the AV	3	0.77	0.54 – 0.72
5. Dominated by the AV	3	0.70	0.49 – 0.57

Table 6. EAVE-Q factor correlation matrix

	Benefits of adherence	Compassionate AV	Turning away from others	Externalising the AV	Dominated by the AV	Total Score[a]
Benefits of adherence	1.000	.568**	.277**	.095	.291**	.731**
Compassionate AV	.568**	1.000	.389**	.243*	.209*	.795**
Turning away from others	.277**	.389**	1.000	.201*	.297**	.582**
Externalising the AV	.095	.243*	.201*	1.000	.121	.451**
Dominated by the AV	.291**	.209*	.297**	.121	1.000	.400**

***** *Significant at the p <.01 level; ** Significant at the p <.001 level*
[a] *Correlation between subscales and EAVE-Q total score*

Table 7. Correlations between EAVE-Q domains and clinical outcome measures

	EAVE-Q total score		**Benefits of adherence**		**Compassionate AV**		**Turning away from others**		**Externalising the AV**		**Dominated by the AV**	
	r_s	**p value**	**r_s**	**p value**	**r_s**	**p value**	r_s	p value	**r_s**	**p value**	**r_s**	**p value**
BMI (n = 148)	.028	(.732)	.059	(.473)	.116	(.161)	-.071	(.390)	-.045	(.588)	-.137	(.097)
Age of onset (n = 148)	-0.86	(.301)	.123	(.135)	-.106	(.199)	**-.313**	**(<.001*)**	-0.99	(.231)	0.11	(.891)
Years since 1st onset (n = 148)	-1.09	(.188)	-0.55	(.507)	-.164	(.046*)	-.079	(.341)	-.017	(.842)	.095	(.249)
EDE-Q Global Mean (n = 148)	**.419**	**(<.001*)**	**.234**	**(<.01*)**	**.312**	**(<.001*)**	**.442**	**(<.001*)**	-.007	(.935)	**.482**	**(<.001*)**
EDE-Q Restraint	**.339**	**(<.001*)**	**.240**	**(<.01*)**	.198	(.016*)	**.322**	**(<.001*)**	-.029	(.727)	**.493**	**(<.001*)**
EDE-Q Eating concern	.159	(.054)	-.015	(.852)	.112	(.176)	**.284**	**(<.001*)**	.022	(.792)	**.314**	**(<.001*)**
EDE-AQ Shape concern	**.401**	**(<.001*)**	**.231**	**(<.01*)**	**.319**	**(<.001*)**	**.410**	**(<.001*)**	-.022	(.791)	**.350**	**(<.001*)**
EDE-Q Weight concern	**.397**	**(<.001*)**	**.237**	**(<.01*)**	**.359**	**(<.001*)**	**.360**	**(<.001*)**	-.042	(.608)	**.350**	**(<.001*)**
EDE-Q Binge eating	.074	(.369)	-0.64	(.442)	.134	(.106)	.140	(.089)	.028	(.740)	.039	(.637)
EDE-Q Vomiting	.079	(.342)	-.010	(.906)	.076	(.360)	.091	(.269)	-.041	(.623)	.171	(.038*)
EDE-Q Laxatives	-.009	(.912)	-.070	(.399)	-.014	(.869)	.125	(.131)	-.101	(.222)	.109	(.187)
EDE-Q Compulsive exercise	.112	(.173)	.042	(.609)	.030	(.721)	.006	(.944)	.110	(.183)	**.226**	**(<.01*)**

	EAVE-Q total score		Benefits of adherence		Compassionate AV		Turning away from others		Externalising the AV		Dominated by the AV	
	r_s	p value	r_s	p value	r_s	p value	r_s	p value	r_s	p value	r_s	p value
DASS-21 domains (n = 141):												
Total score	.246	(.003*)	-.040	(.638)	.150	(.077)	**.472**	**(<.001*)**	.031	(.718)	.268	(.001*)
Depression	.142	(.094)	-.099	(.243)	.074	(.383)	**.396**	**(<.001*)**	-.010	(.904)	.215	(.010*)
Anxiety	.275	(.001*)	.023	(.782)	.187	(.027*)	**.472**	**(<.001*)**	.061	(.472)	.177	(.036*)
Stress	.222	(.008*)	.004	(.966)	.141	(.153)	**.347**	**(<.001*)**	.032	(.706)	**.294**	**(<.001*)**
WHOQOL-BREF Domains (n = 142):												
Physical	-.100	(.238)	.154	(.068)	-.017	(.839)	**-.320**	**(<.001*)**	-.022	(.796)	**-.295**	**(<.001*)**
Psychological	-.148	(.079)	.117	(.166)	-.115	(.173)	**-.377**	**(<.001*)**	.038	(.653)	**-.266**	**(.001*)**
Social	-.177	(.035*)	-.037	(.665)	-.117	(.167)	-.227	**(<.01*)**	.064	(.452)	**-.305**	**(<.001*)**
Environmental	-.053	(.535)	.196	(.019*)	.003	(.976)	**-.302**	**(<.001*)**	.007	(.931)	-.180	(.032*)

*NB As some variables were not normally distributed, all correlations reported are for Spearman's correlation coefficient (r_s) for ease of interpretation, as a sensitivity analysis showed no significant difference in the size and direction of correlations or significance values when using Pearson's r to calculate the associations between normally distributed variables; * denotes significant result; All moderate correlations are highlighted in bold for ease of interpretation.*

Table 8. Comparison of scores between AN-R and AN-BP as assessed using the SCID

EAVE-Q Domain: Mean (SD)	AN-R (n = 36)	AN-BP (n = 14)	Test statistic (df)	p value	MD	95% CI
Total EAVE-Q score	3.33 (0.68)	3.33 (0.55)	-0.15 (48)	.988	.003	-0.41 – 0.41
Benefits of adherence	3.49 (1.00)	3.60 (0.85)	-0.35 (48)	.729	-.106	-0.71 - 0.50
The compassionate AV	2.71 (1.21)	2.96 (1.03)	-0.70 (48)	.489	-.256	-0.99 - 0.48
Turning away from others	3.63 (1.15)	3.88 (0.76)	-0.76 (48)	.454	-.251	-0.92 - 0.42
Externalising the AV	2.94 (1.10)	2.38 (1.14)	1.59 (48)	.120	.554	-0.15 – 1.26
Dominated by the AV	3.98 (0.80)	3.79 (0.93)	0.74 (48)	.460	.196	-0.33 - 0.72

Table 9. Comparison of EAVE-Q scores between a 'clinical' and 'subthreshold' group

EAVE-Q Domain: Mean (SD)	Clinical (n = 104)	Subthreshold (n = 44)	Test statistic (df)	p value	MD	95% CI
Total EAVE-Q score	3.31 (0.68)	3.34 (0.54)	t (146) = 0.32	.748	0.04	-0.19 – 0.27
Benefits of adherence	3.39 (0.97)	3.50 (0.89)	t (146) = 0.64	.526	0.11	-0.23 – 0.44
The compassionate AV	2.80 (1.16)	3.09 (0.98)	t (146) = 1.44	.151	0.29	-0.11 – 0.68
Turning away from others	3.85 (1.04)	3.82 (0.90)	t (146) = -0.19	.849	-0.03	-0.39 – 0.32
Externalising the AV	2.86 (1.09)	2.75 (1.06)	t (146) = -0.56	.577	-0.11	-0.49 – 0.28
Dominated by the AV	3.73 (0.91)	3.54 (0.98)	t (146) = -1.17	.243	-0.20	-0.53 – 0.13

Table 10. Assessment of test-retest reliability using Intraclass Correlation Coefficient (n = 49)

	Completed Time 1 and 2, Mean delay: 29 days (range 6 - 76), n = 49			
Item	**Cronbach's Alpha**	**ICC***	**95% Confidence Interval**	**p value**
Total EAVE-Q score	.85	.74	.58 - .85	<.001
Benefits of adherence	.81	.68	.49 - .80	<.001
The compassionate AV	.78	.64	.44 - .78	<.001
Turning away from others	.87	.76	.62 - .86	<.001
Externalising the AV	.82	.70	.52 - .82	<.001
Dominated by the AV	.82	.68	.50 - .81	<.001

*ICC values using two way mixed effects model, single measurement, absolute agreement.